Seif, S., Mignan, A., Zechar, J., Werner, M., & Wiemer, S. (2017). Estimating ETAS: the effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449-469. https://doi.org/10.1002/2016JB012809

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE

10.1002/2016JB012809

Key Points:
- Missing aftershocks bias the productivity parameters and Omori's c and p for cutoff magnitudes $M < 3.5$
- Missing aftershocks cause greater uncertainty in these parameter estimates
- Branching ratio estimators are biased under realistic conditions with a finite catalog length

Supporting Information:
- Supporting Information S1

Correspondence to:
S. Seif,
stefanie.seif@sed.ethz.ch

Citation:
Seif, S., A. Mignan, J. D. Zechar, M. J. Werner, and S. Wiemer (2017), Estimating ETAS: The effects of truncation, missing data, and model assumptions, *J. Geophys. Res. Solid Earth*, *121*, doi:10.1002/2016JB012809.

Received 8 JAN 2016
Accepted 7 DEC 2016
Accepted article online 9 DEC 2016

Estimating ETAS: The effects of truncation, missing data, and model assumptions

Stefanie Seif[1], Arnaud Mignan[2], Jeremy Douglas Zechar[1], Maximilian Jonas Werner[3], and Stefan Wiemer[1]

[1]Swiss Seismological Service, ETH Zürich, Zürich, Switzerland, [2]Institute of Geophysics, ETH Zürich, Zurich, Switzerland, [3]School of Earth Sciences, University of Bristol, Bristol, UK

Abstract The Epidemic-Type Aftershock Sequence (ETAS) model is widely used to describe the occurrence of earthquakes in space and time, but there has been little discussion dedicated to the limits of, and influences on, its estimation. Among the possible influences we emphasize in this article the effect of the cutoff magnitude, M_{cut}, above which parameters are estimated; the finite length of earthquake catalogs; and missing data (e.g., during lively aftershock sequences). We analyze catalogs from Southern California and Italy and find that some parameters vary as a function of M_{cut} due to changing sample size (which affects, e.g., Omori's c constant) or an intrinsic dependence on M_{cut} (as M_{cut} increases, absolute productivity and background rate decrease). We also explore the influence of another form of truncation—the finite catalog length—that can bias estimators of the branching ratio. Being also a function of Omori's p value, the true branching ratio is underestimated by 45% to 5% for $1.05 < p < 1.2$. Finite sample size affects the variation of the branching ratio estimates. Moreover, we investigate the effect of missing aftershocks and find that the ETAS productivity parameters (α and K_0) and the Omori's c and p values are significantly changed for $M_{cut} < 3.5$. We further find that conventional estimation errors for these parameters, inferred from simulations that do not account for aftershock incompleteness, are underestimated by, on average, a factor of 8.

1. Introduction

The Epidemic-Type Aftershock Sequence (ETAS) model is one of the most widely used statistical models to describe the temporal (and later on spatial) clustering of seismicity [*Ogata*, 1988, 1998]. While other models to describe seismicity have been proposed over the years [*Vere-Jones*, 1970, 1978, 2005; *Kagan and Knopoff*, 1987; *Console and Murru*, 2001; *Lippiello et al*., 2007; *Turcotte et al*., 2007], we focus on ETAS because it is commonly used for research and is being applied in operational earthquake forecasting [*Marzocchi and Lombardi*, 2009; *Marzocchi and Murru*, 2012; *Marzocchi et al*., 2014; *Rhoades et al*., 2015] and time-dependent seismic hazard [*Gerstenberger et al*., 2014; *Field et al*., 2015].

As its name suggests, ETAS was designed to describe aftershocks, in particular by expressing their rate as a function of time (temporal model) or as a function of time and space (spatiotemporal model). ETAS reflects the fundamental observation that aftershocks tend to cluster near and after a main shock. The model consists of two parts: (1) background events, which occur independently but are often interpreted as being caused by the same underlying process (e.g., tectonic loading due to plate movement) and (2) aftershocks, which are triggered by other earthquakes, either background events or previous aftershocks. The independent and triggered parts are described by the conditional intensity function λ which quantifies the earthquake rate at time t and location (x, y) (in case of a spatiotemporal model)

$$\lambda\ (t,x,y|H_t)\ = \mu(x,y) + \sum_{j:t_j<t} g\Big(t-t_j\ ,x-x_j,y-y_j;M_j\Big) \tag{1}$$

where $\mu(x, y)$ [time^{-1} · distance^{-2}] is the spatially heterogeneous background rate and $g(t - t_j,\ x - x_j,\ y - y_j;\ M_j)$ [time^{-1} · distance^{-2}] is the triggering function. The triggering function expresses the influence of past earthquakes with $t_j < t$ or equivalently the history H_t of the seismicity up to time t. It consists of empirical distribution functions, which model the productivity of aftershocks by a parent earthquake and the distribution of aftershocks in time and space. It has the following form [*Ogata*, 1998]:

$$g(t,x,y;M) = K_0\, e^{\alpha(M-M_{\text{cut}})} \cdot c^{p-1}(t+c)^{-p}(p-1) \cdot f(x,y|M) \ , \tag{2}$$

where the temporal and spatial distributions are normalized. In the common ETAS formulation the spatial decay f follows a power law with

$$f(x,y|M) = \left(d\ e^{\gamma\,(M-M_{\text{cut}})}\right)^{q-1}/\pi\cdot\left(x^2+y^2+d\,e^{\gamma\,\left(M_j-M_{\text{cut}}\right)}\right)^{-q}(q-1), \tag{3}$$

which shows good performance in *Ogata and Zhuang* [2006] and *Werner et al*. [2011].

The triggering function provides information about the stability of the earthquake-generating process. The branching ratio n describes a process as stable if $n<1$, critical if $n=1$, and supercritical or explosive if $n>1$. In nature we observe extended earthquake sequences over long periods with stable features, but they could include temporarily explosive processes that culminate in large earthquakes [*Sornette and Helmstetter*, 2002] and could explain accelerated seismicity in the time before large earthquakes (see review by *Mignan* [2011]). n describes the average number of aftershocks per earthquake when averaged over all magnitudes with

$$n = \int_{M_0}^{M_{\max}} K_0 e^{\alpha(M-M_0)}\cdot p(M)\ \mathrm{d}M = \begin{cases} \dfrac{K_0\beta\ 1-e^{-(\beta-\alpha)(M_{\max}-M_0)}}{\beta-\alpha\ \ 1-e^{-\beta(M_{\max}-M_0)}} & \text{for } \alpha\neq\beta \\ \dfrac{K_0\beta(M_{\max}-M_0)}{1-e^{-\beta(M_{\max}-M_0)}} & \text{for } \alpha=\beta \end{cases}, \tag{4}$$

where $p(M)$ is the probability density function (pdf) of the magnitude distribution, $\beta = \log(10)*b$ the exponent of the magnitude distribution, M_0 is the magnitude above which data are assumed to be complete, and $M_{\max}$ is the maximum magnitude assumed in $p(M)$. There exist more definitions of the branching ratio [*Helmstetter and Sornette*, 2003] which are biased under certain conditions, and we explore this bias in section 4.

Readers not familiar with the details of the ETAS model, we refer to the supporting information, where we extensively review the functional form of ETAS: we characterize and explain each parameter of the spatiotemporal ETAS model and briefly summarize the maximum likelihood estimation (MLE) procedure, through which parameter estimates are obtained. We indicate model assumptions and their potential consequences.

The ETAS model is extensively used throughout the literature for various purposes, including: (i) the description and quantification of seismic characteristics of different tectonic zones, with the distribution of the aftershocks in space and time. Hence, ETAS serves as a tool to understand the interplay between tectonic forces and observed seismicity rates [*Kagan et al*., 2010; *Chu et al*., 2011]. (ii) The ETAS model plays a key role in null-hypothesis testing, e.g., about the physics of earthquake generation [*Ogata and Zhuang*, 2006]. Since the ETAS model incorporates the stochasticity of the clustering process, it represents a better null hypothesis than the Poisson hypothesis [*Werner*, 2008] (note that an alternative null hypothesis consists in using the negative binomial distribution to represent earthquake clustering [e.g., *Mignan et al*., 2013], (iii) ETAS is commonly used to prospectively forecast seismicity such as in the Collaboratory for the Study of Earthquake Predictability project (e.g., *Lombardi and Marzocchi* [2010] (Italy); [*Werner et al*. [2011] (California); and *Zhuang* [2011] (Japan)) or in a modified form to forecast induced seismicity [*Bachmann et al*., 2011].

The ETAS model has been developed and applied for almost three decades and is now being used routinely by the statistical seismology community. However, its parameter estimates are influenced by the model formulation and its deviation from the observations. The choice of the model depends on the scientific question and can be formulated, e.g., with or without time-dependent background rate, anisotropic aftershock triggering, spatially and temporally varying ETAS parameters, finite duration of aftershock triggering, and 3-D description of location. Furthermore, parameter estimates are influenced by data truncation through the choice of the magnitude cutoff M_{cut}. We dedicate the present study to several of these points: we discuss the bias introduced by incomplete aftershocks and finite catalog length and the effect of data truncation through the choice of M_{cut}.

The data set can be truncated at any M_{cut} between the minimum magnitude M_c at which the data are thought to be complete (see review and tutorial on M_c by *Mignan and Woessner* [2012]) and the maximum

magnitude above which the ETAS estimation procedure is unstable due to sample size limitations. How this choice affects parameter estimates is discussed in several studies. *Wang et al*. [2010a] found that by increasing the magnitude threshold of the auxiliary window the branching ratio and p decrease, d increases, and α stays approximately the same. *Schoenberg et al*. [2010] found that by increasing M_{cut} in the target window c, d, p, and q change and that the true values of p and q are not recovered. They also observe that the trends of the changes depend on the formulation of the ETAS model. We find changes in the parameter estimates of c and d, too, which is counterintuitive at first, as the time and space distributions are independent of magnitude. In addition to the study of *Schoenberg et al*. [2010] we provide in section 3 an explanation to these observations and analytical formulations for the intrinsic dependence of the background rate and K_0 on M_{cut}. Our argumentations are similar to those of *Harte* [2015] who gives detailed explanations on the origin of parameter bias based on the concept of broken and false linkages. Beyond the effects considered by *Harte* [2015] we investigate the influence of incomplete aftershock sequences on parameter estimates.

Incomplete aftershock sequences occur when M_{cut} is below the completeness threshold of aftershock sequences which is, immediately after a large earthquake, higher than the threshold of general seismicity ($M_{c(af)} > M_{c(bulk)}$). Incomplete data influence the estimates of the productivity parameters α and K_0 [*Helmstetter et al*., 2006; *Werner et al*., 2011; *Hainzl et al*., 2013; *Omi et al*., 2014] and the estimate of c [*Utsu et al*., 1995; *Kagan*, 2004; *Hainzl*, 2016a]. The missing data due to incomplete aftershock sequences represent a substantial fraction of the total earthquake data: *Kagan* [2004] found that following the Landers earthquake in 1992, between 17,500 and 28,000 events above $M = 2$ were missing, which corresponds to roughly 25% of the Southern Californian catalog [*Hauksson et al*., 2012]. An obvious solution to avoid data incompleteness would be to choose a sufficiently large M_{cut}. However, the price for neglecting small earthquakes is large: small earthquakes may provide insight into the seismic process [*Ebel*, 2008; *Mignan*, 2014] or reveal seismicity characteristics more clearly, like the occurrence of foreshocks before large earthquakes [*Mignan*, 2012a; *Schurr et al*., 2014]. Moreover, including small earthquakes reduce the bias of ETAS parameter estimates related to sample size [*Wang et al*., 2010b] and the bias related to the dependence of the parameter estimates on M_{cut} [*Schoenberg et al*., 2010; *Harte*, 2015].

The idea that aftershock incompleteness affects ETAS estimation is accepted among researchers, and there are various approaches to handle the problem. *Kagan* [1991] and *Hainzl et al*. [2008] remove all aftershocks in a certain time window after the main shock. Alternatively, *Helmstetter et al*. [2006] and *Werner et al*. [2011] describe the completeness magnitude as a function of the main shock magnitude and time. But their approach is limited to a different method of estimating parameters: they optimize the parameters for a 1 day forecast and the estimates depend on the length of the forecasting window. *Omi et al*. [2013, 2014] model aftershock incompleteness for each individual aftershock sequence, rather than for a whole earthquake catalog, but they only use the temporal ETAS formulation.

Despite these approaches, little research has been performed to quantify and review the bias of ETAS parameters that arises from missing data. The aim of this article is to investigate the dependence of ETAS parameters on M_{cut} and specifically how missing data in aftershock sequences influence the parameter estimates at different M_{cut}. We also consider how sample size and temporal edge effects can bias the branching ratio, which characterizes the epidemic nature of the process. In section 2, we describe data used to illustrate the effects of M_{cut} and missing aftershocks. Using these data, we explore in section 3 how the ETAS parameter estimates depend on M_{cut}; we also quantify the bias that arises from incomplete aftershock sequences, and we show that the uncertainties of the parameter estimates α, K_0, p, and c may be grossly underestimated when ignoring the effect of missing data. In section 4, we quantify the bias of branching ratio estimators caused by temporal edge effects.

2. Data

To most clearly see how ETAS parameters depend on M_{cut} and missing aftershocks, we investigate earthquake catalogs from Southern California, Italy, and simulations.

2.1. Southern California and Italy Earthquake Catalogs

We use relocated earthquake data from the Southern California catalog between 1981 and 2014 [*Hauksson et al*., 2012] within a polygon that describes the boundary of the Southern California Seismic Network

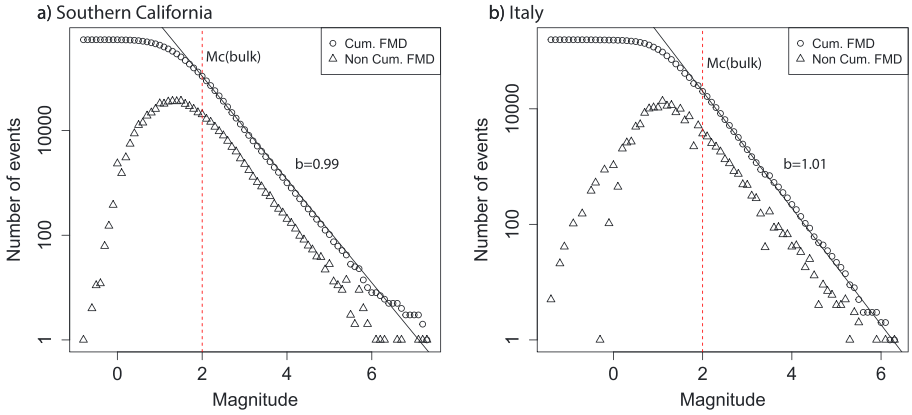

Figure 1. Analysis of the completeness magnitude $M_{c(\text{bulk})}$ determined from the full catalog for (a) Southern California (1981–2014) and (b) Italy (2005–2015).

coverage area [*Hutton et al*., 2010]. For Italy, we analyze the homogenized earthquake catalog by *Gasperini et al*. [2013] which contains calibrated earthquake data with respect to the moment magnitude for the years 1981–2015 within a polygon that covers the mainland of Italy. For the ETAS parameter estimation in Southern California we take the first 5 years and data outside the spatial polygon as supporting information. In Italy we restrict the data to a depth of 30 km. Because of a network change in 2005 and therefore an increased detectability of small earthquakes we consider only data from 16 April 2005 to 31 December 2015 for the MLE and take 1 year as well as data outside the spatial polygon as supporting information.

The magnitude of completeness is an important constraint in any ETAS study because MLE assumes complete data. We estimate $M_{c(\text{bulk})}$ from the full data sets as the magnitude at which the magnitude distribution departs from the G-R law (see review by *Mignan and Woessner* [2012]), more precisely here using the method of *Amorese* [2007]. For Southern California and Italy we find $M_{c(\text{bulk})} = 2$ with standard deviation SD = 0.1 (obtained by bootstrapping) in each case. The MLE of the b value of the G-R law is found $b = 0.99$ and 1.01 for Southern California and Italy, respectively, with binned magnitudes $\Delta M = 0.1$ (Figures 1a and 1b). M_c varies in time especially during large aftershock sequences where the detection threshold is highest in the early stage of aftershock sequences [*Kagan*, 2004; *Helmstetter et al*., 2006; *Ogata and Katsura*, 2006; *Werner et al*., 2011; *Omi et al*., 2013, 2014; *Hainzl*, 2016b]. The value of $\max(M_c(t))$ strongly depends on the estimation method. *Hainzl* [2016b] found for sample sizes of 10 a maximum of approximately $M_c = 4.5$ for the Landers and Hector Mine sequences. Other studies found $max(M_c\ (t)) \approx 4$–4.5 with different methods [*Kagan*, 2004; *Helmstetter et al*., 2006]. However, the true maximum value, immediately after the main shock, cannot be estimated due to a finite resolution but is likely higher than 4.0–4.5. To visualize the variation we compute $M_c(t)$ in Figures 2a and 2b with a sliding window approach. The estimated $M_c(t)$ is highly sensitive to the choice of the window length which determines the degree of smoothing and using 200 earthquakes per window and an offset of 50 earthquakes we find $\max(M_c(t)) = M_{c(af)} = 3.5$ for Southern California. In the case of Italy (Figure 2b), we observe a lower $M_c(t)$ starting from April 2005, which can be related to changes in the seismic network [*Marzocchi and Lombardi*, 2009]. Note that M_c varies in space too [*Mignan*, 2012b]; however, this is not considered here. In the present study we investigate the role of M_{cut} with $M_{c(\text{bulk})} \leq M_{\text{cut}} \lesssim M_{c(af)}$ on ETAS parameter estimation and do not consider the extreme case $M_{\text{cut}} < M_{c(\text{bulk})}$, which would require an in-depth investigation of the intrinsic scale of the earthquake detection process [*Mignan*, 2012b; *Mignan and Chen*, 2015].

2.2. Synthetic Earthquake Catalogs

To explore ETAS estimation in the case where we know the generating model, we consider catalogs from simulations that mimic Southern California. We generate catalogs following the procedure proposed by *Zhuang* [2011] (Figure S1). First, we model the background events. We assume that the number is Poisson distributed with the mean corresponding to the identified background events from the Southern California earthquake catalog using a stochastic declustering method [*Zhuang et al*., 2002], with $M_{\text{cut}} = 2$. We specify their spatial distribution with their smoothed locations using a Gaussian kernel with a bandwidth of 10 km. Simulated background events are distributed according to the resulting spatial density. The

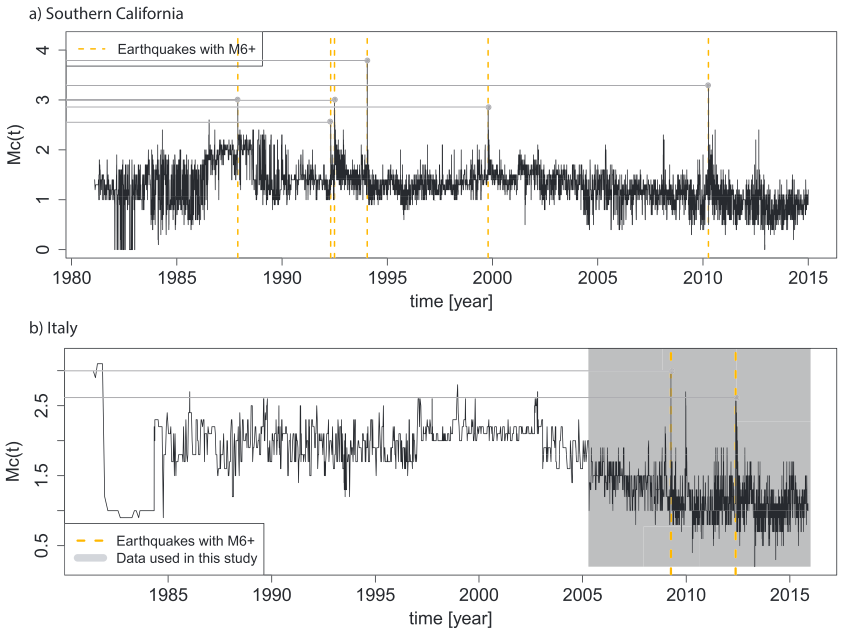

Figure 2. Temporal variation $M_c(t)$ of the completeness magnitude for (a) Southern California (1981–2014) and (b) Italy (1981–2015) calculated using 200 earthquakes per window and an offset of 50 earthquakes.

occurrence times of the background events are sampled from a uniform distribution $U(t_0, t_{max})$, with $t_0 = 0$ and $t_{max} = 33$ years (corresponding to the length of the Southern California catalog), as the earthquake rate follows a stationary Poisson process, and their magnitudes are drawn from the G-R distribution with $b = 0.99$, truncated at $M_{max} = 7.5$ (equation (S4)). We choose M_{max} such that the value represents the maximum observed magnitude, which is $M = 7.3$ (plus an error of 0.2) in the catalog. We choose this value not to represent the maximum magnitude which would be expected to happen in Southern California but to create synthetics which are close to the observed catalog. Once the background events are simulated, we model their aftershocks, assuming that the number of triggered aftershocks is Poisson distributed with the mean rate controlled by the productivity. The magnitudes are sampled from the G-R distribution truncated at $M_{max} = 7.5$ (equation (S4)), the occurrence times from the modified Omori law (equation (2)), and the locations from the isotropic spatial distribution function f (equation (3)). For second-, third-, and higher-generation aftershocks, the triggering step is repeated for every earthquake. We refer the reader to *Zhuang and Touati* [2015] for technical details of the simulation procedure and a simulation code in R. The simulation ends when there are no more potential parents (case $n < 1$; see equation (4)). In the simulation procedure, magnitudes, occurrence times, and locations are stochastic, which makes each realization unique.

Following this procedure, we generate 30 synthetic ETAS catalogs with $M_{cut} = 2$ using ETAS parameters (Table 1) which are estimated following the iterative approach of *Zhuang et al*. [2002] where ETAS parameters and the background rate are obtained simultaneously. We require that $\alpha = \beta$ to correct for the assumption of an isotropic aftershock distribution or temporally stationary background events (see Text S1.2.1). We do not use the exact estimated value for K_0 because it describes a supercritical process. The supercritical branching ratio is probably caused by the underestimation of p (see section 3.1 and *Harte* [2015]). The underestimation of p leads to an overestimation of K_0 because of correlation [*Schoenberg et al*., 2010; *Harte*, 2015]. The overestimation of K_0 increases the value of n, because $n \sim K_0$ (equation (4)). Because simulations require $n < 1$ we adjust K_0 such that $n = 0.99$. We only keep simulations with a number of $M6+$ main shocks and a seismicity rate that are similar to the observed catalog: at least 6 $M6+$ main shocks (8 observed), which guarantees a similar fraction of missing aftershocks (next paragraph) and less than 180,000 events (130,000 observed) due to computational limitations.

To investigate the effect of aftershock incompleteness on estimation, we also use a set of modified simulated catalogs: from each of the original 30 simulations, we remove the aftershocks of $M5+$ earthquakes. In particular, we remove aftershocks below the threshold $M_c(t, m) = m - 4.5 - 0.75 \cdot \log_{10}(t)$ following equation

Table 1. ETAS Parameters Used for Creating Synthetic Catalogs With $M_{cut} = 2$ and $\alpha = \beta = 2.29$[a]

Parameter	$\alpha = \beta = 2.29$
K_0	0.089 (0.080)
α	2.29
c (day)	0.011
p	1.08
d (km^2)	0.0019
q	1.47
γ	2.01
bg-rate (events/yr)	1070
n	0.99

[a]The parameters are estimated from Southern California seismicity (1981–2014, with 5 years history). K_0 in parentheses is adjusted to obtain a stable process ($n = 0.99$).

(15) of *Helmstetter et al*. [2006] derived from Southern California where m is the main shock magnitude and t is the time since it occurred. We apply the removal of aftershocks only to the descendants (first or higher order) of the main shock to assure not to remove spatially unrelated earthquakes. With this procedure we remove about 10% of the data in each catalog. In what follows we refer to these catalogs as complete simulated catalogs and incomplete simulated catalogs, respectively.

3. Influence of Cutoff Magnitude on ETAS Parameter Estimates

In every application of the ETAS model, the modeler must choose a minimum magnitude of interest, M_{cut}. It is important that M_{cut} is larger than M_c, at which the data are thought to be complete because ETAS does not account for missing data. It is also well known that in the wake of a large event, aftershocks that would normally be detected are missed, that is $M_{c(\text{bulk})} < M_{c(af)}$ (section 2). Data with magnitudes above $M_{c(af)}$ should be complete and a direct application of the MLE method should be possible. However, to maximize the sample size, one usually chooses for M_{cut} an estimate of $M_{c(\text{bulk})}$, knowing that this underestimates the completeness magnitude in the presence of large aftershock sequences. Here we investigate the effect of the M_{cut} for complete catalogs and data incompleteness for $M_{c(\text{bulk})} \leq M_{cut} \lesssim M_{c(af)}$ for incomplete catalogs on the ETAS parameter estimates.

We estimate ETAS parameters using the MLE method (see Text S1.3) for Southern California, Italy, and synthetic catalogs setting M_{cut} to each value in the set {2.0, 2.5, 3.0, 3.5, 4.0, 4.5}. For Southern California and Italy we estimate the parameters both with and without the condition that $\alpha = \beta$; for the simulated catalogs we do only the latter. The dependence of the parameter estimates on M_{cut} for Southern California and Italy is shown in Figure 3 and listed in Table 2. The parameter estimates obtained from the two sets of synthetic catalogs are compared in Figure 4. We also compared the estimated parameters pairwise, to be sure that any variation is not due to randomness in the limited data set of 30 catalogs. In the following sections we describe the M_{cut} dependence of each parameter and compare the estimates from the observations of Southern California with the simulations.

3.1. K_0 and p

K_0 describes the mean number of offspring above M_{cut} of a parent with the magnitude M_{cut}. In other words K_0 intrinsically depends on M_{cut}: K_0 stays constant for $\alpha = \beta$ and decreases for $\alpha \neq \beta$ with M_{cut} (Figure 3, K_0 and inset). The dependence of K_0 on M_{cut} can be understood with the following scenario: Considering $M_{cut1} < M_{cut2} < M_{obs}$, we expect the number of earthquakes above M_{obs} (denoted as N_1 and N_2 for the corresponding cutoff magnitudes), generated by a main shock with magnitude m, to be the same. It holds that

$$\begin{aligned} \frac{N_1}{N_2} &= \frac{K_1}{K_2}\cdot\frac{e^{\alpha(m-M_{cut1})}}{e^{\alpha(m-M_{cut2})}}\cdot\frac{e^{-\beta(M_{obs}-M_{cut1})}}{e^{-\beta(M_{obs}-M_{cut2})}} \\ 1 &= \frac{K_1}{K_2}\cdot e^{(\beta-\alpha)(M_{cut1}-M_{cut2})} \\ K_2 &= \begin{cases} K_1\cdot e^{(\beta-\alpha)(M_{cut1}-M_{cut2})} & \text{for } \alpha \neq \beta \\ K_1 & \text{for } \alpha = \beta \end{cases}. \end{aligned} \tag{5}$$

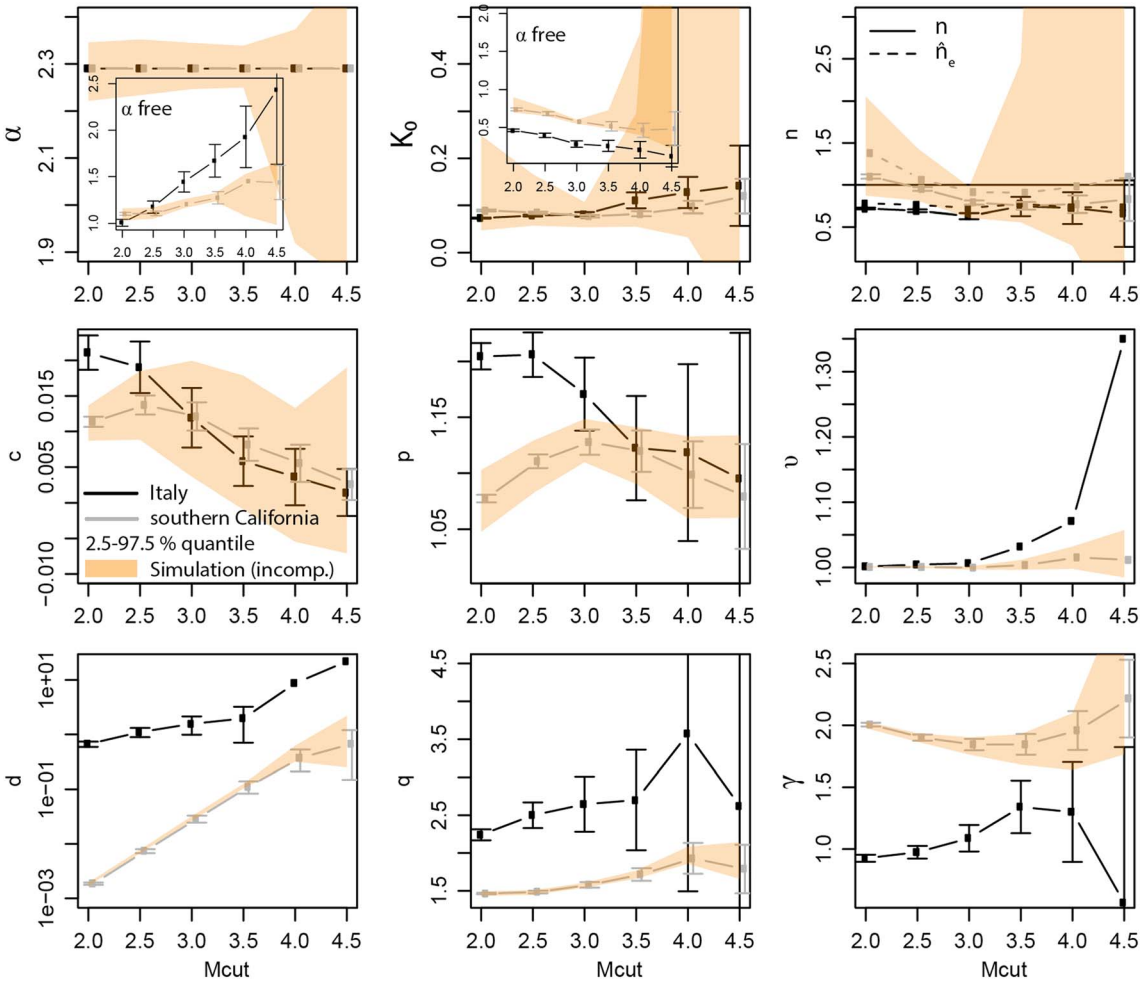

Figure 3. Parameter estimates are plotted against M_{cut} for $\alpha = \beta$. Data: Italy (2005–2015), black curve; and Southern California (1981–2014), gray curve. The vertical bars represent 2 times the standard errors which are determined from the approximated inverse Hessian of the log-likelihood function. For Southern California these errors can be compared with the simulation based 95% quantiles of synthetic incomplete catalogs (orange, compare Figure 4), which are added to the median estimates of the observed catalogs. In the case where α is treated as a free parameter all estimates but α and K_0 follow the same trends as for $\alpha = \beta$. The estimates for α and K_0 are shown in the inlet. In order to be able to interpret the Southern Californian estimates, we restrict the y axis only to its range of estimates.

From equation (5) we expect K_0 to be constant for $\alpha = \beta$. However, its value changes in both reality (Figure 3, K_0) and simulations (Figure 4, K_0). The simulation reveals furthermore that K_0 is overestimated. The overestimation of K_0 may be due to its anticorrelation with the underestimated p (Figure 4, p).

p is the exponent of the temporal power law and describes the decay rate of aftershocks. The temporal power law is magnitude independent, so we would expect a constant p for different M_{cut}. However, we observe a decrease. A decrease of p for large M_{cut} is also reported in *Harte* [2015]. He explained the decrease of p as a consequence of the decreasing sample size which shifts the peak of the mass close to $t = 0$ to higher values. We test this hypothesis with simulations, where we simulate a large sample of the temporal distribution function (equation (2), Omori law) and estimate parameters p and c for randomly thinned subsamples. We observe the contrary of *Harte's* [2015] theory: p stays almost constant with a slight increase with decreasing sample size (Figure 5, p increases $<0.5\%$ at a sample size of 20) and moreover p is not correlated with the smallest waiting time (correlation coefficient = 0.03). In order to find an explanation for the observed contradiction between *Harte's* [2015] theory and our simulations, one fact may play the key role: p is not only determined by the highest density part of the power law around $t = 0$ but also by the large tail of the distribution, which is less sensitive to sample size. Our investigations suggest that for the estimation of p the weight of the tail is larger than the weight of the highest density part.

Table 2. ETAS Parameters Estimated From the Southern Californian Data (1981–2014, With 5 Years History)[a]

	Set 1	Set 2	Set 3	Set 4
Parameter	$M_{cut} = 3.5$ α free	α fixed	$M_{cut} = 2.5$ α free	α fixed
K_0	0.51	0.082	0.68	0.084
α	1.27	2.29	1.10	2.29
c (day)	0.004	0.008	0.007	0.014
p	1.09	1.12	1.09	1.11
d (km^2)	0.231	0.111	0.016	0.007
q	1.59	1.72	1.37	1.90
γ	1.31	1.85	1.35	1.49
bg-rate	38	45	305	434
n	1.13	0.75	1.31	0.95

[a]Parameter Sets 1 to 4 are estimated at $M_{cut} = 3.5$ and $M_{cut} = 2.5$ and the parameter sets are distinguished between $\alpha = \beta$ and α free.

Since the decreasing sample size does not explain the decrease of p, the concept of broken links offers an alternative explanation. There, links between parents with $M < M_{cut}$ and their children are broken and consequently wrongly associated with another parent event. Orphaned events, which should be correctly referred to as background events, are more likely associated as direct aftershocks if they happen in the aftershock tail of a large earthquake. This widens the distribution of direct aftershocks, and it leads to a larger tail, which can be modeled by a smaller p.

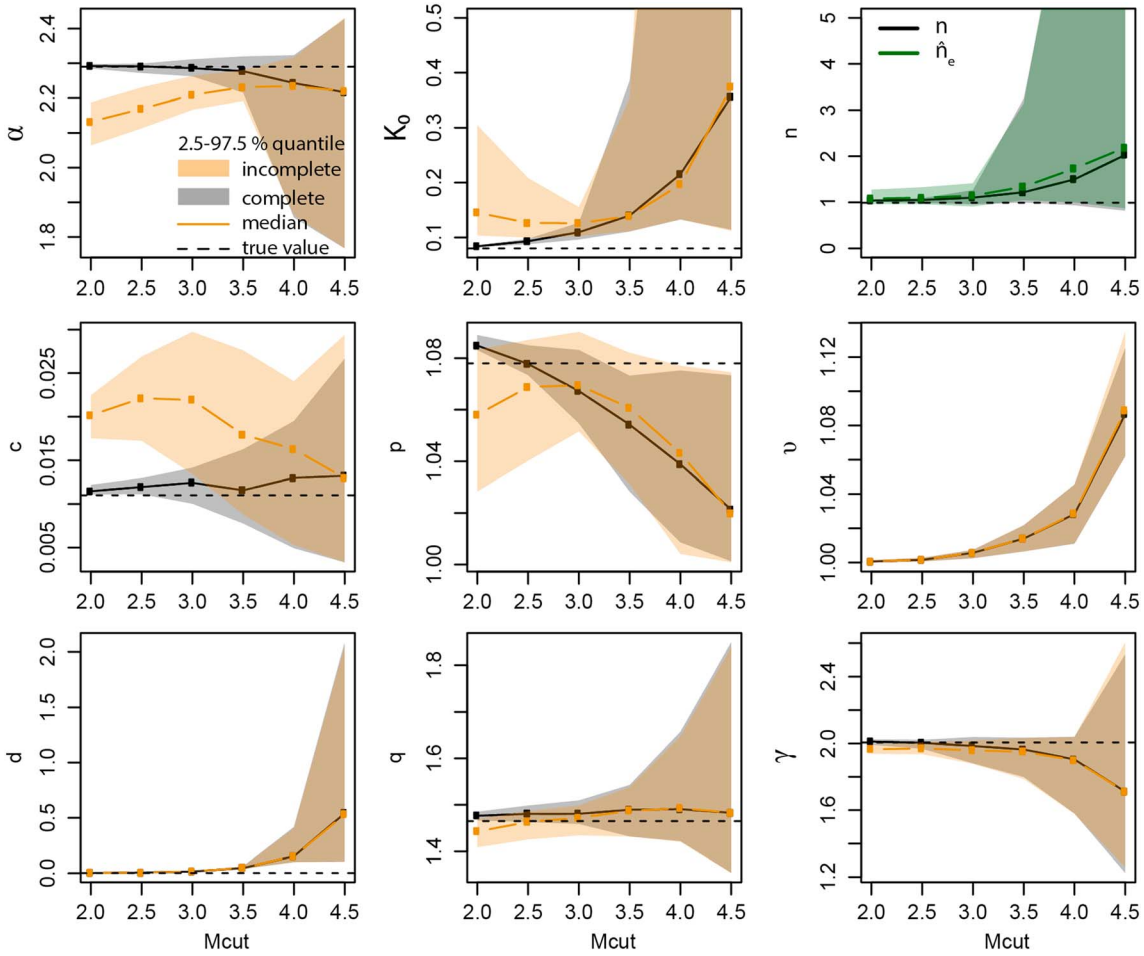

Figure 4. Parameter estimates are plotted against M_{cut} for 30 synthetic earthquake catalogs simulated with parameters from Table 1 (gray) and compared with the parameter estimates where aftershocks are removed (orange). The catalogs are generated as described in section 2.2. The branching ratio n (equation (4)) and $\hat{n}_e$ (equation (9)) are estimates of the complete catalogs.

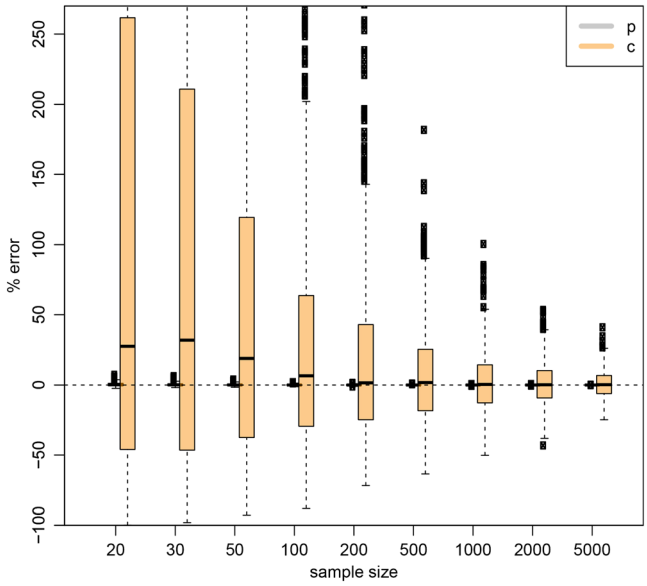

Figure 5. Boxplot of the % error (e.g., $\frac{\hat{c}-c}{c} \cdot 100$) of the estimated $\hat{c}$ and $\hat{p}$ for different sample sizes. We simulated a power law distribution in time according to equation (2) (Omori law) with a sample size of $n = 5000$ and estimated parameters c and p for $n = 5000$ and randomly thinned samples. We repeated the simulations 1000 times. The median c increases with decreasing sample size and p is not strongly affected by the sample size.

Although p and K_0 converge to the true value for decreasing M_{cut} there remains a small bias (overestimation) at M_0. An overestimation of p was also reported in *Schoenberg et al*. [2010]. We suspect that the overestimation depends on the model formulation and perhaps on integration problems of the spatial density function, which is a power law function in our study and in *Schoenberg et al*. [2010], that cannot be integrated analytically over an arbitrary polygon as pointed out by *Harte* [2015]. Furthermore, the bias of K_0 and p at M_0 may be explained by the poor performance of the MLE procedure in a close to critical regime ($n \sim 1$), because we find no bias for $n = 0.32$.

We observe that both K_0 and p are affected by incomplete aftershock sequences. K_0 is overestimated by 70% at $M_{cut} = 2$ and p is underestimated by 2% with respect to the corresponding mean estimate of the complete catalog. The variation in the parameter estimates, represented by the 95% confidence interval from the simulated catalogs, for the incomplete catalogs is about 50 (10) times larger for K_0 (p) at $M_{cut} = 2$ (Table 3).

3.2. α

From the incomplete simulated catalog (Figure 4, α orange) we see that α is significantly underestimated for small $M_{cut} \leq 3$ at a significance level of 0.05 (the 95% confidence bounds between complete and incomplete estimates do not overlap) and its value increases with M_{cut} toward its true value. The variation in the parameter estimates is larger for the incomplete than for the complete catalogs. Instead of decreasing with M_{cut}, as expected by *Schoenberg et al*. [2010] and *Wang et al*. [2010b], the uncertainty is almost constant and at the smallest $M_{cut} = 2$ is 10 times as large as the conventionally calculated variation from a complete data set (Table 3).

In the real catalog where α is a free parameter, the α value decreases by 32% from $M_{cut} = 4$ to 2 (see Figure 3, α inset). In the synthetic incomplete catalog α decreases by only 3% in the same M_{cut} range. Missing data due to aftershock incompleteness can therefore not explain the observed decrease of α with decreasing M_{cut}. An alternative explanation for the decrease, which is not captured in the synthetic catalogs, is an anisotropic aftershock distribution, which leads to a low α (S1.2.1) [*Hainzl et al*., 2008] or nonstationary background rates, which are shown to have significant influence on α in Southern California [*Hainzl et al*., 2013]. *Hainzl et al*. [2013] also show that the intensity of the bias of α increases with a smaller c value. The bias of α in reality, where c is likely lower than estimated (see section 3.3), is hence expected to be larger than we find.

Table 3. Parameter Variation of Incomplete Data Divided by Complete Data[a]

M_{cut}	K_0	α	c	p	d	q	γ	n	υ
2	51.3	9.7	4.5	9.5	1.4	3.3	1.6	25	1.1
2.5	11.1	4.3	5.3	4.0	1.2	1.7	1.4	6	1.0
3	1.7	2.0	4.0	1.4	1.1	1.3	0.9	1.2	1.0
3.5	0.9	0.9	2.2	1.1	1.0	1.0	1.0	0.8	1.0
4	1.7	1.0	1.3	1.1	1.0	1.0	1.0	1.7	1.0
4.5	1.2	1.0	1.1	1.0	1.0	1.0	1.0	1.5	1.1

[a]1.0 means no differences in variation between incomplete and complete data.

3.3. *c*

We observe from the estimates of the real and simulated catalogs that c depends on M_{cut} (Figures 3 and 4, *c*). To determine whether the dependence arises from the varying sample size, which decreases with increasing M_{cut}, we sample a temporal distribution according to the Omori law in equation (2), with different sample sizes and estimate the relevant parameters c and p (Figure 5). The median c value increases about 30% at a sample size of 20. This can be understood because with decreasing sample size, aftershock sequences are randomly thinned which reduces the effective mass at times around 0 and this can be modeled by increasing c [*Harte*, 2015]. We find further evidence for this explanation in the large correlation between c and the smallest waiting time (correlation coefficient = 0.91). However, we cannot exclude the influence of further sources, e.g., broken links between parents below M_{cut} and children, on the c value.

Incomplete simulated catalogs (Figure 4, *c* orange) reveal that aftershock incompleteness increases the c value, where the increase is more prominent in the middle range of M_{cut} (for $M_{cut} = 2.5$, the medians from the incomplete and complete simulations differ by 85%). An increase in the c value due to aftershock incompleteness was also reported by *Utsu et al*. [1995], *Kagan and Houston* [2005], and *Hainzl* [2016a]. Both the influence of the sample size and the aftershock incompleteness on c suggest that c depends on the magnitude M of the main shock, because it determines the sample size through the number of aftershocks, and the degree of incompleteness. However, both influences suggest a contrary dependence of c on M: sample size suggests a positive and aftershock incompleteness a negative correlation.

3.4. *d, q,* γ

d is the spatial equivalent to c, and following the same reasoning as described in section 3.3, it depends on the sample size and therefore increases with M_{cut}. A different formulation of the spatial distribution f would probably lead to a different effect [*Schoenberg et al*., 2010]. d is not affected by incomplete aftershocks (Figure 4, *d*). q, γ have a slightly increasing/decreasing trend, with increasing $M_{cut,}$ respectively (Figure 4, *q*, γ). The increase of q can be explained by the decreasing sample size (Figure 5). q does not decrease with M_{cut} as its equivalent p (Figure 4, *p*, *q*), due to false links. In space secondary aftershocks distribute isotropically and not one sided as in time; and hence, a wrong association with a parent event leads to a diffusion-type widening of the distribution which is weaker. q estimated from observations shows the same trend as the simulations and γ shows variations with no clear trend (Figure 3, q and γ). The parameter estimates seem to be affected by another source, not present in the simulations, which could be anisotropic aftershock distribution [*Helmstetter et al*., 2005; *Hainzl et al*., 2008] or, more generally, inhomogeneous or inconsistent data.

3.5. *n*

The true and empirical branching ratios (equation (4) and equation (4) where $p(M)$ is replaced by the empirical magnitude distribution) yield supercritical values for small M_{cut} when considering the whole Southern California seismicity (1981–2014) and for all M_{cut} when considering the whole Italian seismicity (1981–2015) (not shown in the paper). These supercritical branching ratios are not realistic for extensive catalogs and could be related to data or estimation problems. Data problems include (a) uncertainty of the earthquake source parameters, (b) heterogeneity of seismicity recordings, and (c) incomplete data. Investigating point (a) in Southern California resulted in no significant difference in n after parameter reestimation with perturbed magnitudes. Investigating (b), we found that the declustered seismicity is not stationary from 1981 to 2014. After taking only data from 2001 to 2014 (using 5 years of data as history before this period) we still obtain

$n > 1$ for $M_{cut} = 2$ which could be explained by (c) since $M_{c(bulk)} < M_c(x,y,t)$ locally (see section 2). Finally, biases in the parameter estimates, especially in K_0 and p, could lead to supercritical branching ratios (see sections 3.1 and 5). p is shown to be underestimated at low M_{cut} due to aftershock incompleteness and at large M_{cut} probably due to false links (see Figure 4, p and section 3.1). Its underestimation leads to an overestimation of K_0 because of their negative correlation. This leads to an overestimation of n because of the positive correlation between K_0 and n. Further research on the branching ratio would be necessary to specify the origin of $n > 1$.

3.6. Background Rate $\rho(x,y)$ and Relaxation Parameter υ

It is trivial that the background rate $\nu \cdot \rho(x,y)$ [events $\cdot$ area^{-1} $\cdot$ time^{-1}] at any point (x, y) decreases with increasing M_{cut}. But the scaling of ETAS background event rate does not follow the G-R law. This is because, in any application, earthquakes that are triggered by events with magnitude less than M_{cut} will be identified as apparent background events if the minimum triggering magnitude $M_0 < M_{cut}$ [*Sornette and Werner*, 2005]. Therefore, the background rate estimated from data with $M_0 < M_{cut}$ is only an apparent background rate, and likewise the branching ratio is an apparent branching ratio. Using the apparent branching ratio, n_a^{obs}, defined in *Sornette and Werner* [2005] we derive the number of apparent background events at a given M_{cut} (see Appendix A) as

$$N_{bg}^{cut} = N_{tot}^{cut} \cdot \left(1 - n_a^{obs} \cdot \frac{e^{(\beta-\alpha)(M_{max}-M_{cut})} - 1}{e^{(\beta-\alpha)(M_{max}-M_{obs})} - 1}\right)$$
$$= N_{tot}^{obs} \cdot e^{\beta \cdot (M_{obs}-M_{cut})} \cdot \left(1 - \frac{N_{tot}^{obs} - N_{bg}^{obs}}{N_{tot}^{obs}} \cdot \frac{e^{(\beta-\alpha)(M_{max}-M_{cut})} - 1}{e^{(\beta-\alpha)(M_{max}-M_{obs})} - 1}\right). \quad (6)$$

In Figure 6 we compare the true number of background events, obtained through the known triggering history of ETAS simulations, at varying M_{cut} with different estimates thereof: estimates based on equation (6), based on the G-R law, and based on the declustering method of *Zaliapin et al*. [2008]. We find that equation (6) (knowing the N_{bg}^{obs} and N_{tot}^{obs} at $M_{obs} = 3.5$) represents the true background rate well (i.e., within the standard deviation confidence bounds). The estimate based on the G-R relation overpredicts at $M_{cut} < 3.5$ (knowing N_{bg}^{obs} at $M_{obs} = 3.5$). The Zaliapin declustering slightly underestimates at $M_{cut} > 2$ and overestimates at small $M_{cut} = 2$.

υ is the relaxation parameter or weighting constant for the background rate, also called failure rate in *Console and Murru* [2001]. It is introduced to assure that the integral of the conditional intensity function (equation (1)) over the target spatiotemporal window equals the number of target earthquakes and is hence not a free parameter. Therefore, the integral of $\nu \cdot \rho(x,y)$ over the target area should equal the number of background events N_{bg}. It is $\nu = N_{bg}/\int_{x,y}\rho(x,y)dxdy$. When using stochastic declustering, N_{bg} is the sum over all background probabilities. υ increases with M_{cut} for the simulated catalogs (Figure 4, υ) and the true observations (Figure 3, υ).

4. Bias of Branching Ratio Estimators

The branching ratio is not an ETAS parameter but rather an important property of the process; it has an unambiguous interpretation and implications for the criticality, and therefore predictability, of earthquakes. The branching ratio has the following definitions [*Helmstetter and Sornette*, 2003]: (i) the average number of aftershocks per earthquake when averaged over all magnitudes, (ii) the fraction of events that are triggered, or (iii) the fraction of aftershocks that are second generation or higher generation. (i) is expressed in equation (4) and (ii) and (iii) are expressed, respectively, as

$$\hat{n} = \frac{N_{af}}{N_{tot}}, \text{ and} \quad (7)$$

$$\hat{n} = \frac{N_{af}^{2nd+higher}}{N_{af}}, \quad (8)$$

where N_{af} is the number of aftershocks and N_{tot} the total number of earthquakes. In practice, we can replace

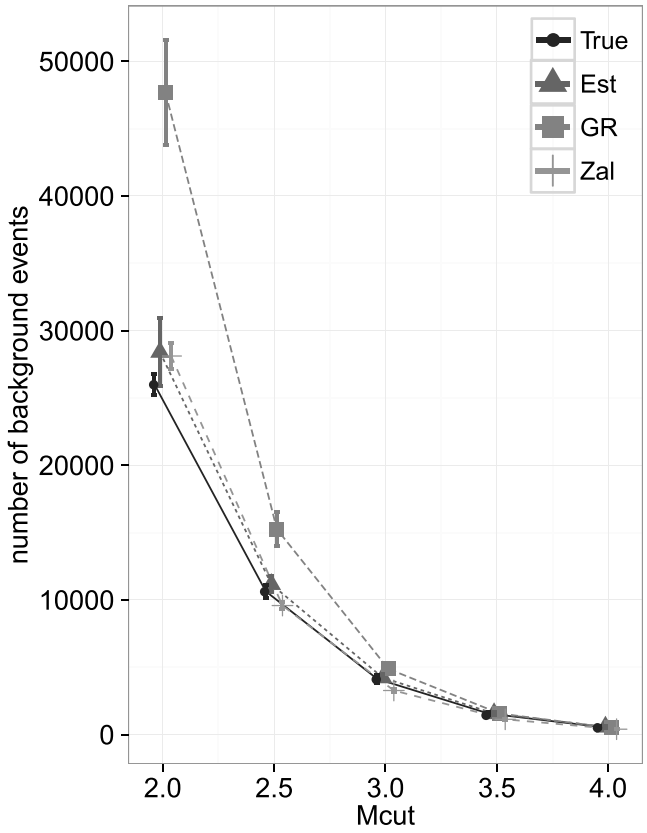

Figure 6. Number of background events (True) of 30 synthetic catalog generated with parameters from Table 1, which are obtained from the triggering history, is compared with the estimated number of background events (Est) obtained with equation (15), given N_{bg}^{obs} and N_{tot}^{obs} at $M_{obs} = 3.5$. N_{bg}^{obs} is the exact number of background events at M_{obs}. True is also compared with the Gutenberg-Richter (G-R) prediction using N_{bg}^{obs} and with the number of background events obtained by the *Zaliapin et al*. [2008] declustering (Zal).

$p(M)$ in equation (4) with the empirical magnitude distribution, avoiding the stochasticity of the magnitude distribution, and this yields

$$\hat{n}_e = \sum_{i=1}^{N} K_0 e^{\alpha(M_i - M_0)}/N \tag{9}$$

(J. Zhuang, personal communication, 2015). Equation (4) is the true branching ratio of an ETAS process, if one knows the true value of every relevant parameter (K_0, α, β, M_0, M_{max}), while equations (7–9) are estimators that one applies to real seismicity catalogs via declustering or applies to simulated catalogs, where the triggering history (i.e., all connections between parents and children) is known. Equation (4) also becomes an estimator in case it is calculated with parameter estimates.

The estimators of the branching ratio in equations 7–9 have never been explored for their consistency. We want to explore the catalogs' temporal edge effects and finite sample properties to detect their influence on possible bias and variation of the branching ratio estimators in equations 7–9. To focus only on the bias introduced by limited sample size and finite duration of the catalog, we assume that N_{af} and N_{tot} are known through the triggering history of the simulations, such that we can rule out any bias that is induced from other unmet model assumptions, e.g., finite aftershock duration. By temporal edge effects, we mean effects arising from the fact that any catalog we analyze (i) will include events triggered by earthquakes occurring before the beginning of the catalog and (ii) will miss events triggered after the end of the catalog. We investigate the influence of the finite sample size as a function of the productivity parameters and the memory length, determined by the size of p and the number of events in a catalog. Those were also found to be sources of influence in a related investigation, carried out by *Sornette and Utkin* [2009], who evaluated the performance of different declustering methods by comparing $\hat{n}$ (equation (7)) obtained from declustering

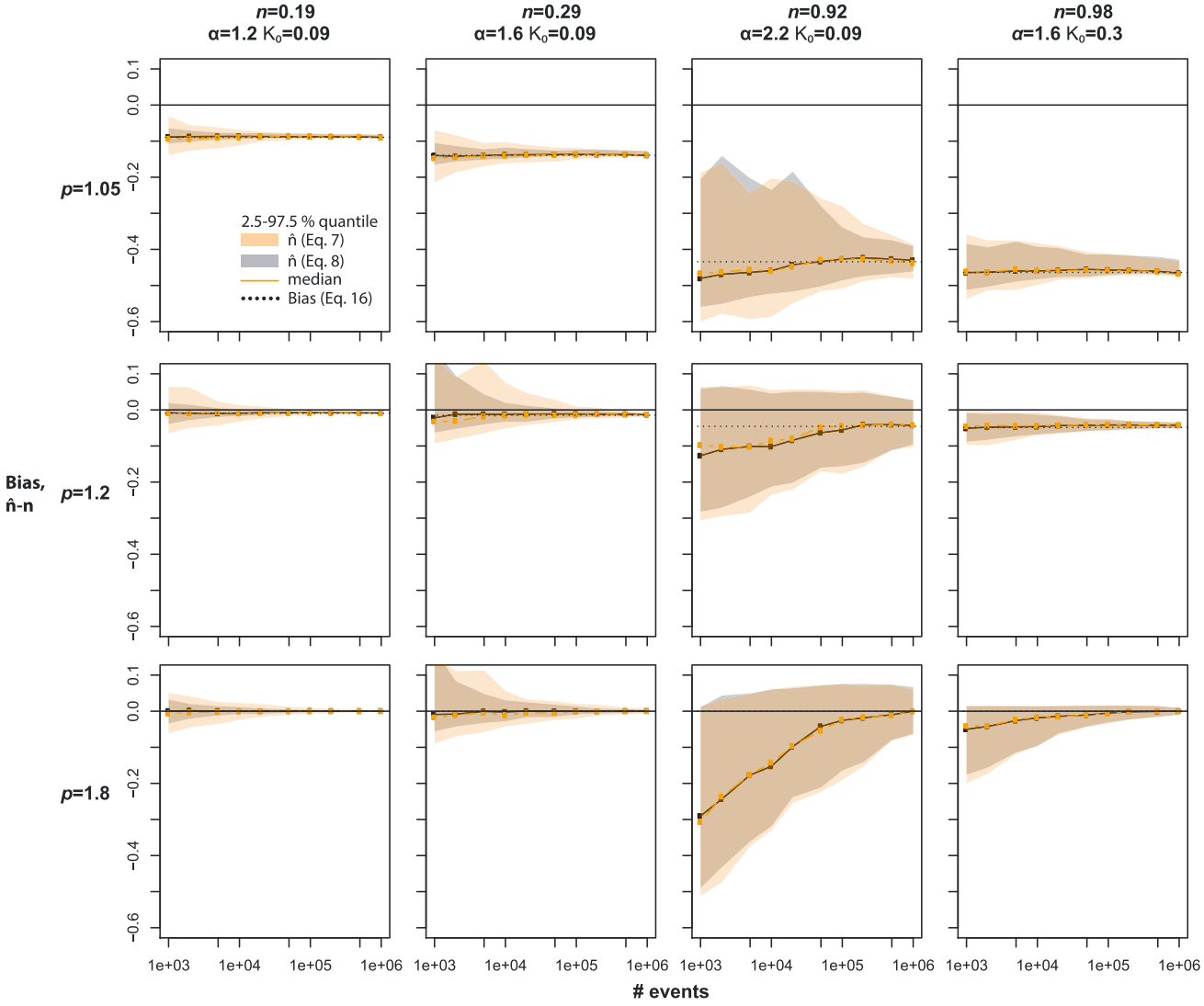

Figure 7. Bias of the observed branching ratios $\hat{n}$ (equations (7) and (8)) from the true n (equation (4)). $\hat{n}$ is determined for different (first–fourth columns) productivity parameters, (first–third rows) p, and number of events (x axis). One hundred synthetic catalogs are simulated (parameters from Table 1) over 100 years with minimum 1,000,000 events in total.

techniques with $\hat{n}$ from synthetic catalogs. For various values of α, K_0, and p, we generate 100 synthetic ETAS catalogs with at least 1 million events and apply each of the estimators. Note that because we vary the productivity parameters, we also implicitly vary the true branching ratio. To consider how the estimates vary as a function of sample size, we subsample the 100 synthetic catalogs, starting at the end of the catalog and growing backward. For example, to estimate bias from catalogs containing 1000 events, we cull the final 1000 events of each catalog; we start from the end of the catalog so we have the most history available.

In Figures 7 and 8 we show the bias and spread of the estimators of equations (7) and (8) and equation (9), as a function of sample size for different α, K_0, and p. We find that the variation of the estimates in equations (7) and (8) increases as α increases (compare the shaded regions in the first three columns of any row in Figure 7) and decreases with sample size (Figure 7, any plot). The variation of both estimates is almost the same but slightly larger for $\hat{n}$ of equation (8) (Figure 7, gray versus orange shaded regions). Under realistic conditions ($1.05 < p < 1.2$; Figure 7, first and second rows) $\hat{n}$ does not converge to the true n. The size of the bias depends on p (different rows in Figure 7). If p is unrealistically large (Figure 7, third row), equations (7) and (8) seem to be consistent estimators, and when combining a large p with a small α (Figure 7, third row and first and second columns), the bias is small even for small sample sizes.

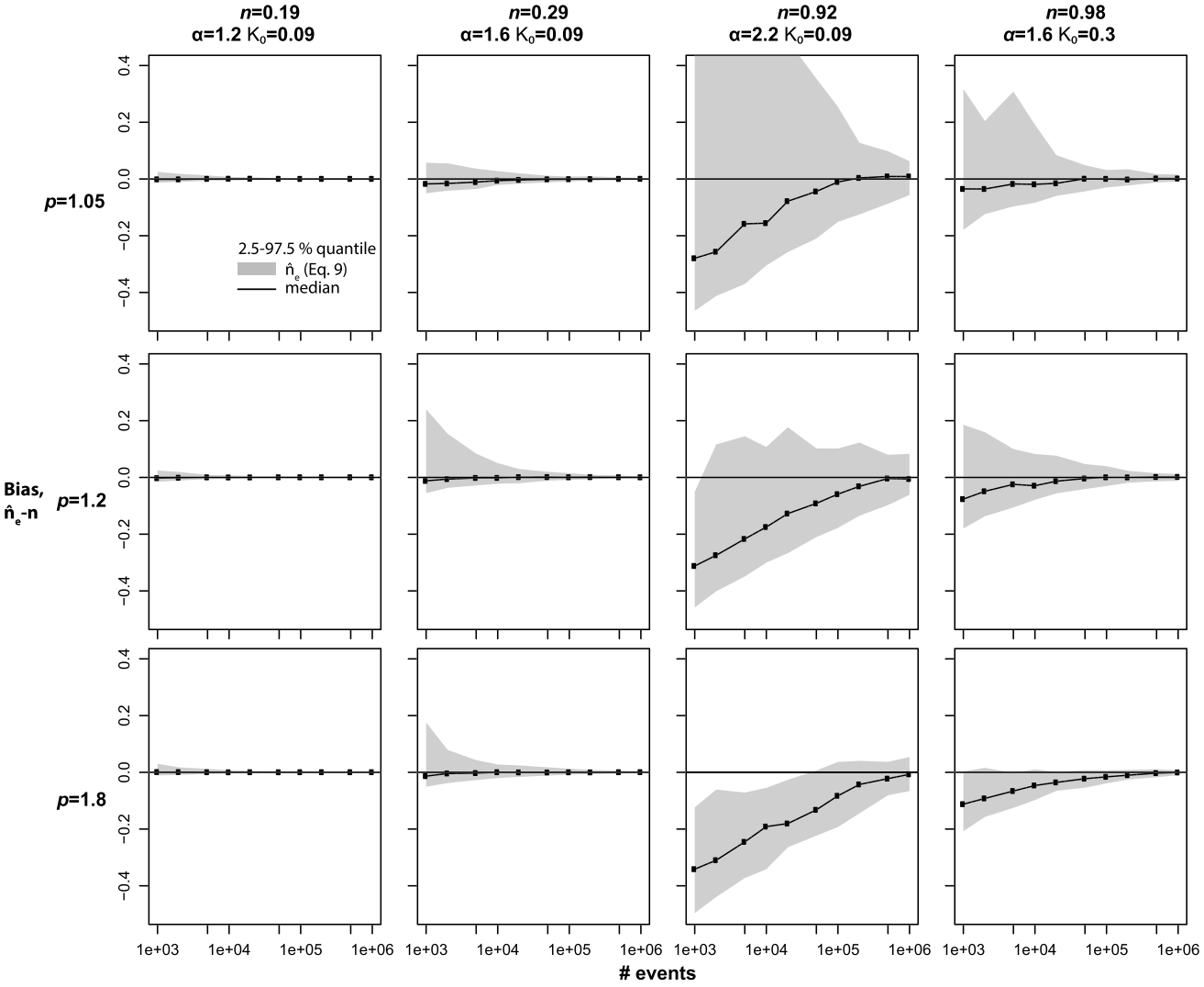

Figure 8. Bias of the empirical branching ratio $\hat{n}_e$ from the true n (equations (9) and (4)). $\hat{n}_e$ is determined for different (first–fourth columns) productivity parameters, (first–third rows) p, and number of events (x axis). One hundred synthetic catalogs are simulated (parameters from Table 1) over 100 years with minimum 1,000,000 events in total.

We intend to show now that the observed bias is caused by temporal edge effects and the effect of p on the memory length. Consider a catalog divided into an auxiliary window lasting from t_0 to t_1, a target window from t_1 to t_2, and a posttarget window from t_2 to ∞, with $t_0 < t_1 < t_2$. Equations (7) and (8) are calculated using events from the target window, and assume we know the complete triggering history, so we know if each event in the target window is triggered or background. Certainly, the target window contains some events triggered by parents in the auxiliary window—call these incoming aftershocks—and some aftershocks from parents in the target window will fall in the posttarget window—call these outgoing aftershocks. For an event that occurs in the target window at time t_x, we denote by $f_{\text{out}}(t_x)$ the fraction of its aftershocks that are outgoing. It is described by integration of the modified Omori law (pdf)

$$g(t - t_x) = \frac{p-1}{c}\left(1 + \frac{t - t_x}{c}\right)^{-p}, \tag{10}$$

with

$$f_{\text{out}}(t_x) = \int_{t_2}^{\infty} g(t - t_x)\mathrm{d}t = \left(1 + \frac{t_2 - t_x}{c}\right)^{1-p}. \tag{11}$$

The average fraction of outgoing aftershocks per time (normalized by the length of the target window) is

$$F_{\text{out}} = \int_{t_1}^{t_2} \frac{f_{\text{out}}(t_x)\mathrm{d}t_x}{t_2 - t_1} = \frac{c}{-2+p}\left(1 - \left(\frac{c - t_1 + t_2}{c}\right)^{2-p}\right)\frac{1}{t_2 - t_1}. \tag{12}$$

For an event that occurs at t_y in the auxiliary window, the fraction of its aftershocks that are incoming is

$$f_{\text{in}}(t_y) = \int_{t_1}^{t_2} g(t - t_y)\mathrm{d}t. \tag{13}$$

Likewise, the average fraction of incoming aftershocks per time (normalized by the length of the target window) is

$$\begin{aligned} F_{\text{in}} &= \int_{t_0}^{t_1} f_{\text{in}}(t_y)\ \ \mathrm{d}t_y\ \frac{1}{t_2 - t_1} \\ &= \frac{c}{-2+p}\left(1 - \left(\frac{c - t_0 + t_1}{c}\right)^{2-p} + \left(\frac{c - t_0 + t_2}{c}\right)^{2-p} - \left(\frac{c - t_1 + t_2}{c}\right)^{2-p}\right)\frac{1}{t_2 - t_1}. \end{aligned} \tag{14}$$

If $F_{\text{in}} = F_{\text{out}}$, the branching ratio estimators are unbiased. In reality, with finite catalog lengths, the equation is not balanced, resulting in

$$F_{\text{bias}}(p, c, t_0, t_1,\ t_2) = F_{\text{out}} - F_{\text{in}} = \frac{c^{p-1}}{-2+p}\left((c + t_1 - t_0)^{2-p} - (c + t_2 - t_0)^{2-p}\right)/(t_2 - t_1), \tag{15}$$

and the bias is expressed with

$$\hat{n} - n = -n \cdot F_{\text{bias}}. \tag{16}$$

For $t_2 \to \infty$ equation (15) converges (very slowly) to 0, which means that for realistic conditions with finite t_2 the incoming aftershocks cannot compensate for the outgoing aftershocks. The effect is stronger for smaller p, because the decay rate of aftershocks is lower and hence the memory longer. In Figure 7 we plot $-n \cdot F_{\text{bias}}$ for the approximate conditions at a sample size of 0.5 million events: $t_0 = 0$, $t_1 = 50$, and $t_2 = 100$ years and we observe that it describes the deviation of the estimators from the true value well. When using one of the branching ratio estimators (equations (7) and (8)) one should be aware of the bias and we recommend correcting them with equation (16). We expect the same dependence of the bias on q in the spatial domain. However, the value of q has a larger variability (Table S1), and for large q the bias should be smaller.

For small sample sizes $\hat{n}_e$ consistently underestimates n, and this bias increases with n and α (Figure 8, different columns in the same row). The reason is that for large α most aftershocks are produced by large events (see S1.2.1), so $\hat{n}_e$ strongly depends on the frequency and the exact magnitude of large events, which are subject to the stochasticity of the G-R law and not completely sampled (especially for large magnitudes) for finite processes. The bias of $\hat{n}_e$ decreases for larger catalogs and converges toward 0 (Figure 8, x axis within one plot)—in other words, it appears to be a consistent estimator despite temporal edge effects.

5. Discussion

We find that ETAS parameter estimates depend on M_{cut}, and we detect several reasons for this dependence. Sample size changes the clustering that is observed: as M_{cut} increases, aftershock clusters are randomly thinned and the intensity of their clustering decreases. The effect of this thinning is primarily seen in the parameters c and d of the temporal and spatial power laws, respectively, and to a smaller extent in q which increases for increasing M_{cut} (Figure 4, c, d, and q). The dependence of c on the sample size argues for a main shock magnitude dependent $c(M)$, because the aftershock productivity is dictated by M. The exponent q is less sensitive to sample size because it is determined by the decay of the heavy tail of the distribution, and the median changes by a factor less than 0.05 (Figure 4, q).

The sample size also affects the background rate: it decreases with higher M_{cut} not according to the G-R law but according to equation (6) which accounts for apparent background seismicity [*Sornette and Werner*, 2005]

and which is also found by *Harte* [2015] who additionally investigated second-order biases. K_0 directly depends on M_{cut} for $\alpha \neq \beta$ and follows equation (5).

p shows the opposite behavior of its spatial equivalent: it decreases with M_{cut}. Its behavior can be explained by broken links between parents with $M < M_{cut}$ and their children. Rather than being assigned as background events those orphaned children are more likely assigned as direct aftershocks if they happen in the tail of an aftershock sequence. This widens the distribution of direct aftershocks, which can be modeled by a smaller p.

The same trends for c and d have also been observed for *Schoenberg et al*. [2010] model 2 ($\gamma = 0$); however, for model 1 ($\gamma = \alpha$) the trend for d is reverse, which indicates that the trends might be subject to the ETAS formulation. They observe a bias for p and q at the lowest M_{cut}, which we also observe for p. We think that in our case the overestimation could result from the model formulation or poor performance of the MLE in a close to critical regime. Parameters α and K_0 are not specifically discussed in *Schoenberg et al*. [2010]. Our findings agree with *Wang et al*. [2010a]: they find a decrease of p and an increase of d with increasing M_{cut}, α stays constant. They do not find significant variations for the remaining parameters. *Harte* [2015] analyzes the bias in the parameter estimates introduced by M_{cut} more theoretically. He finds the same trends for c, d, p, K_0, and the background rate and expresses the dependence of the background rate and K_0 similarly to us but using a different theoretical concept, namely on broken linkages between parents and children. In agreement to our results he finds a strong dependence of p on M_{cut}, which he, in contrast to our findings, leads back to the sample size. His p and K_0 estimates are unbiased at M_0, whereas ours are slightly overestimated (p overestimation also observed in *Schoenberg et al*. [2010]).

ETAS parameter estimates also depend on M_{cut} because of incomplete data after large earthquakes: this incompleteness affects the estimates of α, K_0, c, and p (Figure 4, orange). We compare the parameter estimates from complete synthetic catalogs with estimates from more realistic incomplete synthetic catalogs. We find that the influence of aftershock incompleteness is not significant for $M_{cut} \geq 3.5$. This can be explained by the decreasing fraction of missing data for increasing M_{cut}: when $M_{cut} = 4$ the data should be almost complete, which is supported by very similar estimates of α, K_0, c, and p from complete and incomplete simulated catalogs. Only for $M_{cut} \leq 3$ ($M_{cut} \leq 2.5$, 2) do we find a significant difference in α (K_0, c, and p) between complete and incomplete catalogs. The median α estimate at $M_{cut} = 2$ from the incomplete catalogs is 7% smaller than from the complete catalogs. A low α reduces the aftershock productivity of a large main shock and describes the incomplete data immediately after the main shock better. Furthermore, K_0 estimated from incomplete catalogs at $M_{cut} = 2$ is 70% larger than the estimate from complete catalogs. The high K_0 compensates for the low α. c at $M_{cut} = 2.5$ ($M_{cut} = 2$) is 85% (75%) larger which supports, e.g., *Kagan*'s [2004], *Hainzl*'s [2016a], and *Mignan*'s [2016b] claim that c could be an artifact of aftershock incompleteness, which reduces the sample size. And because the intensity of the aftershock incompleteness is determined by the main shock magnitude M, it should be considered to model c as a function of M. p is also influenced by incomplete aftershock sequences and is underestimated by 2% at $M_{cut} = 2$.

Within the MLE of incomplete simulated catalogs, α decreases by 3% between $M_{cut} = 4$ and $M_{cut} = 2$ (Figure 4, α orange). This decrease is much smaller than in the observed catalog, where we also have to deal with missing data, and where we observe a decrease of 32% in this magnitude range (Figure 3, α inset). The higher decrease of α in the observed catalog could be due to the anisotropic distribution of aftershocks [*Hainzl et al*., 2008] or time-dependent background rate [*Hainzl et al*., 2013]. They found that α is systematically underestimated when including the spatial component in the inversion or when assuming a temporal stationary background rate. An estimate of the exact value of α is given in other studies which similarly excluded spatial information and found $\alpha \approx \beta$ [*Felzer*, 2004; *Helmstetter et al*., 2005] and furthermore that Bath's law can be reproduced with this assumption [*Felzer*, 2002]. Also, simulations show that assuming a large α describes observed seismicity well, but only when missing data are modeled in the simulations too [*Helmstetter et al*., 2006].

The variations of the parameter estimates for the synthetic catalogs, depicted by the shaded 95% confidence regions in Figure 4, show large differences between complete and incomplete catalogs for α, K_0, c, and p. As expected, in complete catalogs the spread of the estimates, indicated by the width of the shaded region, is smallest when the sample size is largest, i.e., when M_{cut} is smallest [see also *Schoenberg et al*., 2010; *Wang et al*., 2010b]. For the incomplete catalogs, the variations of α, K_0, and p show the opposite trend—they

increase with decreasing M_{cut} where more data are missing. The variation of c does not increase, probably because the influence of the varying sample size due to M_{cut} is larger. However, the variation of c estimated from incomplete catalogs is larger compared to complete catalogs (Table 3). The variation increases because we describe incomplete data with a model that assumes complete data. The intensity of incompleteness is varying for different data sets and depends strongly on the frequency of large ($M6+$) earthquakes and on the magnitude of these large earthquakes. Data sets with many large earthquakes are more incomplete and should result in a lower α than data sets with few large earthquakes. Fitting a complete data set does not have this dependency and hence a lower variation. For the variation and the estimate itself, we expect that their degree of dependence on M_{cut} changes with the chosen data sets and the underlying model (i.e., the choice of the spatial formulation), but trends should remain similar. The proposed analysis could be used to estimate the dependences accordingly.

The fluctuations of the observed parameter estimates with M_{cut} is large and cannot be explained by their uncertainty estimates based on simulations with aftershock incomplete data (Figure 4). This indicates that in reality (1) data inhomogeneities could be present, e.g., the completeness could generally vary in space and time, (2) the assumptions of a point process model (i.e., earthquakes are points in space and time) are not met, and (3) the chosen model formulation is a poor representation of nature that does not account for, e.g., time-dependent background rate, anisotropic triggering, finite duration of triggering, temporal and spatial varying ETAS parameters, or 3-D earthquake locations.

In this study we wish to point out effects on the two-dimensional representation of the ETAS model. However, we acknowledge that the three-dimensional effects require a more complex spatial description of the seismogenic layer which may lead to different parameter estimates [*Guo et al*., 2015]. Thus, further investigation of the three-dimensional space is justified but beyond the scope of this paper.

6. Conclusion

There exist different definitions of the branching ratio, which distinguish between estimates (obtained by single realizations of the point process) and the true branching ratio, obtained by the parameters of the point process which define the mean number of aftershocks per earthquake when averaged over all magnitudes. We found that the properties of the branching ratio estimators $\hat{n}$ (equations (7) and (8)) depend on the sample size and the temporal finiteness of the catalog. Specifically, a large sample size reduces the variation of the estimates, and the temporal finiteness introduces a bias of the estimates, even if the process is sampled sufficiently (1 million events). We found that under realistic conditions, with catalog lengths smaller than several thousand years and p values between 1.05 and 1.2, $\hat{n}$ is inconsistent. We provide an analytical formulation to estimate its bias in equation (16) as a function of catalog length, p and c, under the assumption of a sufficiently sampled catalog. Our findings show that with perfect knowledge of the triggering structure, one is not able to recover the true n. Consequently, declustering methods, which aim to recover the triggering structure, would always suffer from the bias in $\hat{n}$ if they showed perfect performance. We described the temporal decay with the standard modified Omori formulation; for different formulations, e.g., a truncated power law decay [*Hainzl et al*., 2016] or stretched exponential [*Mignan*, 2015, 2016a, 2016b], we expect a different size of the bias.

From the second part of the paper we conclude that ETAS parameter estimates depend on M_{cut}. The sample size affects the clustering behavior observable in the parameters c and d of the temporal and spatial power laws, which increase with increasing M_{cut} (Figure 4, c and d). M_{cut} also affects the background rate (equation (15)) and K_0 for $\alpha \neq \beta$ (equation (14)). We find that ETAS parameter estimates α, K_0, c, and p are affected by incomplete data after large earthquakes. Incomplete data affect the productivity because fewer events are recorded in comparison to complete data. The Omori parameters are affected because incompleteness varies as a function of time. In the presence of incomplete data, α and p are underestimated, while K_0 and c are overestimated. The bias increases for decreasing M_{cut} and is significant for $M_{cut} \leq 3$. This is supported by *Helmstetter et al*. [2006], who found that it is important to correct for the incompleteness when forecasting aftershock seismicity in time.

Correcting for the induced bias that we found in our simulations is difficult because the bias depends on the parameters, and we estimate parameters from incomplete data. Our simulations show that aftershock

incompleteness leads to an underestimated α, and we find this trend mirrored in the estimates from observations. Aftershock incompleteness is thus a third identified source of negative bias of α. The other two are the assumption of isotropy in the spatial aftershock distribution [*Hainzl et al.*, 2008] and a temporally stationary background rate [*Hainzl et al.*, 2013]. Together, the three sources may explain why α values estimated from MLE of the ETAS model are smaller than the values obtained from more direct measures of productivity (e.g., $\alpha=\beta$). Based on our simulations with $\alpha=\beta$, the expected bias from aftershock incompleteness alone is about 10%. Aftershock anisotropy is presumably a greater source of bias. Tectonic setting influences the anisotropy and may thus affect the bias, too. If reasons exist to assume $\alpha=\beta$, such as those listed at the end of section S1.2.1, then α should be fixed during the MLE. In that case, the corresponding K_0 that results from MLE is probably determined by its correlation to α and to a lower degree influenced by the bias. Correcting c and p for the effect of aftershock incompleteness is not straightforward because the bias likely depends on the other parameters of the process.

We also conclude that the standard deviations of α, K_0, c, and p from synthetic incomplete data increase with decreasing M_{cut} and are orders of magnitudes larger than the commonly estimated standard deviations from complete data for $M_{\text{cut}} \leq 3$ (Table 3). Therefore, when applying the commonly used methods to estimate standard deviations the uncertainties of α, K_0, c, and p are underestimated. We recommend to estimate the uncertainties for these parameters at $M_{\text{cut}} \leq 3$ from simulations which incorporate incomplete aftershock sequences.

This study demonstrates that missing data have a significant effect on the likelihood inference of parameters at low $M_{\text{cut}} \leq 3$ and hence on the conclusions that are drawn from the data. We infer that missing data on smaller and more complex scales in time and space may influence the parameter estimates too, and this influence should also be investigated. The sources of influence on parameter estimates investigated in this paper (cutoff magnitude, aftershock incompleteness, finite catalog length) only form a subset of all potential sources of influence, but we show that they have to be considered when drawing conclusions about seismicity from parameter estimates of different regions or different M_{cut}.

Appendix A

We derive the relationship between the number of apparent background events and M_{cut} expressed by equation (6).

Using equation (4) n_a can be expressed with

$$n_a = \int_{M_{\text{cut}}}^{M_{\max}} K_0 e^{\alpha(M-M_0)} \cdot p(M)\ \mathrm{d}M = \begin{cases} n\left(\dfrac{e^{(\beta-\alpha)(M_{\max}-M_{\text{cut}})}-1}{e^{(\beta-\alpha)(M_{\max}-M_0)}-1}\right) & \text{for } \alpha \neq \beta \\ n\left(\dfrac{M_{\max}-M_{\text{cut}}}{M_{\max}-M_0}\right) & \text{for } \alpha = \beta \end{cases}. \tag{A1}$$

The dependence of the apparent background rate on M_{cut} is derived in the following. First, the total number of events at M_{cut} is estimated according to the G-R law.

$$N_{\text{total}}^{\text{cut}} = N_{\text{total}}^{\text{obs}} \cdot e^{\beta(M_{\text{obs}}-M_{\text{cut}})} \tag{A2}$$

where the total number of events at magnitude M_{obs} is known. Then the number of background events is estimated through equation (7) as

$$N_{\text{bg}}^{\text{cut}} = N_{\text{total}}^{\text{cut}} \cdot \left(1 - n_a^{\text{cut}}\right) \tag{A3}$$

where n_a^{cut} is given with equation (A1) as

$$n_a^{\text{cut}} = n_a^{\text{obs}} \left(\frac{e^{(\beta-\alpha)(M_{\max}-M_{\text{cut}})}-1}{e^{(\beta-\alpha)(M_{\max}-M_{\text{obs}})}-1}\right). \tag{A4}$$

Inserting equation (A4) in (A3) the total number of background events is

$$N_{\text{bg}}^{\text{cut}} = N_{\text{total}}^{\text{cut}} \cdot \left(1 - n_a^{\text{obs}} \frac{e^{(\beta-\alpha)(M_{\max}-M_{\text{cut}})} - 1}{e^{(\beta-\alpha)(M_{\max}-M_{\text{obs}})} - 1}\right)$$
$$= N_{\text{total}}^{\text{obs}} e^{\beta(M_{\text{obs}}-M_{\text{cut}})} \cdot \left(1 - \frac{N_{\text{tot}}^{\text{obs}} - N_{bg}^{\text{obs}}}{N_{\text{tot}}^{\text{obs}}} \frac{e^{(\beta-\alpha)(M_{\max}-M_{\text{cut}})} - 1}{e^{(\beta-\alpha)(M_{\max}-M_{\text{obs}})} - 1}\right). \quad \text{(A5)}$$

Acknowledgments

We thank Jiancang Zhuang for providing the code for estimating ETAS parameters and for his assistance with technical questions. We also thank three anonymous reviewers and the Associate Editor for their helpful reviews. The FORTRAN code for estimating ETAS parameters can be found at http://bemlar.ism.ac.jp/zhuang/software.html. We thank Paolo Gasperini, Barbara Lolli, and Gianfranco Vannucci for providing the Italian catalog. Earthquake data for Southern California were obtained from http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html.

References

Amorese, D. (2007), Applying a change-point detection method on frequency-magnitude distributions, *Bull. Seismol. Soc. Am.*, *97*(5), 1742–1749, doi:10.1785/0120060181.

Bachmann, C. E., S. Wiemer, J. Woessner, and S. Hainzl (2011), Statistical analysis of the induced Basel 2006 earthquake sequence: Introducing a probability-based monitoring approach for enhanced geothermal systems, *Geophys. J. Int.*, *186*(2), 793–807, doi:10.1111/j.1365-246X.2011.05068.x.

Chu, A., F. P. Schoenberg, P. Bird, D. D. Jackson, and Y. Y. Kagan (2011), Comparison of ETAS parameter estimates across different global tectonic zones, *Bull. Seismol. Soc. Am.*, *101*(5), 2323–2339, doi:10.1785/0120100115.

Console, R. (2003), Refining earthquake clustering models, *J. Geophys. Res.*, *108*(B10), 2468, doi:10.1029/2002JB002130.

Console, R., and M. Murru (2001), A simple and testable model for earthquake clustering, *J. Geophys. Res.*, *106*, 8699–8711, doi:10.1029/2000JB900269.

Ebel, J. E. (2008), The importance of small earthquakes, *Seismol. Res. Lett.*, *79*(4), 491–493, doi:10.1785/gssrl.79.4.491.

Felzer, K. R. (2002), Triggering of the 1999 M W 7.1 Hector Mine earthquake by aftershocks of the 1992 M_w 7.3 Landers earthquake, *J. Geophys. Res.*, *107*(B9), 2190, doi:10.1029/2001JB000911.

Felzer, K. R. (2004), A common origin for aftershocks, foreshocks, and multiplets, *Bull. Seismol. Soc. Am.*, *94*(1), 88–98, doi:10.1785/0120030069.

Field, E. H., et al. (2015), UCERF3-ETAS including spatiotemporal clustering for a California Operational Earthquake Forecast (OEF), Oral Session Presented at the StatSei 9 Meeting, Potsdam, Germany.

Gasperini, P., B. Lolli, and G. Vannucci (2013), Empirical calibration of local magnitude data sets versus moment magnitude in Italy, *Bull. Seismol. Soc. Am.*, *103*(4), 2227–2246, doi:10.1785/0120120356.

Gerstenberger, M., G. McVerry, D. Rhoades, and M. Stirling (2014), Seismic hazard modeling for the recovery of Christchurch, *Earthquake Spectra*, *30*(1), 17–29, doi:10.1193/021913EQS037M.

Guo, Y., J. Zhuang, and S. Zhou (2015), A hypocentral version of the space–time ETAS model, *Geophys. J. Int.*, *203*(1), 366–372, doi:10.1093/gji/ggv319.

Gutenberg, B., and C. F. Richter (1944), Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.*, *34*, 185–188.

Hainzl, S. (2016a), Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs, *J. Geophys. Res. Solid Earth*, *121*, 6499–6509, doi:10.1002/2016JB013319.

Hainzl, S. (2016b), Rate dependent incompleteness of earthquake catalogs, *Seismol. Res. Lett.*, *87*(2A), 337–344, doi:10.1785/0220150211.

Hainzl, S., and Y. Ogata (2005), Detecting fluid signals in seismicity data through statistical earthquake modeling, *J. Geophys. Res.*, *110*, B05S07, doi:10.1029/2004JB003247.

Hainzl, S., A. Christophersen, and B. Enescu (2008), Impact of earthquake rupture extensions on parameter estimations of point-process models, *Bull. Seismol. Soc. Am.*, *98*(4), 2066–2072, doi:10.1785/0120070256.

Hainzl, S., O. Zakharova, and D. Marsan (2013), Impact of aseismic transients on the estimation of aftershock productivity parameters, *Bull. Seismol. Soc. Am.*, *103*(3), 1723–1732, doi:10.1785/0120120247.

Hainzl, S., A. Christophersen, D. Rhoades, and D. Harte (2016), Statistical estimation of the duration of aftershock sequences, *Geophys. J. Int.*, *205*(2), 1180–1189, doi:10.1093/gji/ggw075.

Harte, D. S. (2015), Model parameter estimation bias induced by earthquake magnitude cut-off, *Geophys. J. Int.*, *204*(2), 1266–1287, doi:10.1093/gji/ggv524.

Hauksson, E., W. Yang, and P. M. Shearer (2012), Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.*, *102*(5), 2239–2244, doi:10.1785/0120120010.

Helmstetter, A., and D. Sornette (2003), Importance of direct and indirect triggered seismicity in the ETAS model of seismicity, *Geophys. Res. Lett.*, *30*(11), 1576, doi:10.1029/2003GL017670.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2005), Importance of small earthquakes for stress transfers and earthquake triggering, *J. Geophys. Res.*, *110*, B05S08, doi:10.1029/2004JB003286.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for Southern California, *Bull. Seismol. Soc. Am.*, *96*(1), 90–106, doi:10.1785/0120050067.

Hutton, K., J. Woessner, and E. Hauksson (2010), Earthquake monitoring in Southern California for seventy-seven years (1932–2008), *Bull. Seismol. Soc. Am.*, *100*(2), 423–446, doi:10.1785/0120090130.

Kagan, Y. Y. (1991), Likelihood analysis of earthquake catalogues, *Geophys. J. Int.*, *106*(1), 135–148, doi:10.1111/j.1365-246X.1991.tb04607.x.

Kagan, Y. Y. (2004), Short-term properties of earthquake catalogs and models of earthquake source, *Bull. Seismol. Soc. Am.*, *94*(4), 1207–1228, doi:10.1785/012003098.

Kagan, Y. Y., and H. Houston (2005), Relation between mainshock rupture process and Omori's law for aftershock moment release rate, *Geophys. J. Int.*, *163*(3), 1039–1048, doi:10.1111/j.1365-246X.2005.02772.x.

Kagan, Y. Y., and D. D. Jackson (2000), Probabilistic forecasting of earthquakes, *Geophys. J. Int.*, *143*(2), 438–453, doi:10.1046/j.1365-246X.2000.01267.x.

Kagan, Y. Y., and L. Knopoff (1987), Statistical short-term earthquake prediction, *Science*, *236*, 1563–1567, doi:10.1126/science.236.4808.1563.

Kagan, Y. Y., P. Bird, and D. D. Jackson (2010), Earthquake patterns in diverse tectonic zones of the globe, *Pure Appl. Geophys.*, *167*(6-7), 721–741, doi:10.1007/s00024-010-0075-3.

Lippiello, E., C. Godano, and L. de Arcangelis (2007), Dynamical scaling in branching models for seismicity, *Phys. Rev. Lett.*, *98*(9), 098501, doi:10.1103/PhysRevLett.98.098501.

Lippiello, E., F. Giacco, L. D. Arcangelis, W. Marzocchi, and C. Godano (2014), Parameter estimation in the ETAS Model: Approximations and novel methods, *Bull. Seismol. Soc. Am.*, *104*(2), 985–994, doi:10.1785/0120130148.

Llenos, A. L., and J. J. McGuire (2011), Detecting aseismic strain transients from seismicity data, *J. Geophys. Res.*, *116*, B06305, doi:10.1029/2010JB007537.
Lombardi, A. M., and W. Marzocchi (2010), The ETAS model for daily forecasting of Italian seismicity in the CSEP experiment, *Ann. Geophys.*, *53*(3), 155–164, doi:10.4401/ag-4848.
Marzocchi, W., and A. M. Lombardi (2009), Real-time forecasting following a damaging earthquake, *Geophys. Res. Lett.*, *36*, L21302, doi:10.1029/2009GL040233.
Marzocchi, W., and M. Murru (2012), Daily earthquake forecasts during the May–June 2012 Emilia earthquake sequence (northern Italy), *Ann. Geophys.*, *55*(4), 561–567, doi:10.4401/ag-6161.
Marzocchi, W., A. M. Lombardi, and E. Casarotti (2014), The establishment of an operational earthquake forecasting system in Italy, *Seismol. Res. Lett.*, *85*(5), 961–969, doi:10.1785/0220130219.
Mignan, A. (2011), Retrospective on the Accelerating Seismic Release (ASR) hypothesis: Controversy and new horizons, *Tectonophysics*, *505*(1-4), 1–16, doi:10.1016/j.tecto.2011.03.010.
Mignan, A. (2012a), Seismicity precursors to large earthquakes unified in a stress accumulation framework, *Geophys. Res. Lett.*, *39*, L21308, doi:10.1029/2012GL053946.
Mignan, A. (2012b), Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude, *J. Geophys. Res.*, *117*, B08302, doi:10.1029/2012JB009347.
Mignan, A. (2014), The debate on the prognostic value of earthquake foreshocks: A meta-analysis, *Sci. Rep.*, *4*, 4099, doi:10.1038/srep04099.
Mignan, A. (2015), Modeling aftershocks as a stretched exponential relaxation, *Geophys. Res. Lett.*, *42*, 9726–9732, doi:10.1002/2015GL066232.
Mignan, A. (2016a), Revisiting the 1894 Omori aftershock dataset with the stretched exponential function, *Seismol. Res. Lett.*, *87*(3), 685–689, doi:10.1785/0220150230.
Mignan, A. (2016b), Reply to "Comment on 'Revisiting the 1894 Omori aftershock dataset with the stretched exponential function' by A. Mignan" by S. Hainzl and A. Christophersen, *Seismol. Res. Lett.*, *87*(5), 1134–1137, doi:10.1785/0220160110.
Mignan, A., and C.-C. Chen (2015), The spatial scale of detected seismicity, *Pure Appl. Geophys.*, *119*, doi:10.1007/s00024-015-1133-7.
Mignan, A., and J. Woessner (2012), Understanding seismicity catalogs and their problems: Estimating the magnitude of completeness for earthquake catalogs, *Community Online Resour. Stat. Seismol. Anal.*, doi:10.5078/corssa-00180805.
Mignan, A., C. Jiang, J. D. Zechar, S. Wiemer, Z. Wu, and Z. Huang (2013), Completeness of the mainland China earthquake catalog and implications for the setup of the China Earthquake Forecast Testing Center, *Bull. Seismol. Soc. Am.*, *103*(2A), 845–859, doi:10.1785/0120120052.
Ogata, Y. (1988), Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, *83*(401), 9–27.
Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.*, *50*(2), 379–402, doi:10.1023/A:1003403601725.
Ogata, Y., and K. Katsura (2006), Immediate and updated forecasting of aftershock hazard, *Geophys. Res. Lett.*, *33*, L10305, doi:10.1029/2006GL025888.
Ogata, Y., and J. Zhuang (2006), Space–time ETAS models and an improved extension, *Tectonophysics*, *413*(1-2), 13–23, doi:10.1016/j.tecto.2005.10.016.
Omi, T., Y. Ogata, Y. Hirata, and K. Aihara (2013), Forecasting large aftershocks within one day after the main shock, *Sci. Rep.*, *3*(2218), 1–7, doi:10.1038/srep02218.
Omi, T., Y. Ogata, Y. Hirata, and K. Aihara (2014), Estimating the ETAS model from an early aftershock sequence, *Geophys. Res. Lett.*, *41*, 850–857, doi:10.1002/2013GL058958.
Rhoades, D. A., M. Liukis, A. Christophersen, and M. C. Gerstenberger (2015), Retrospective tests of hybrid operational earthquake forecasting models for Canterbury, *Geophys. J. Int.*, *204*(1), 440–456, doi:10.1093/gji/ggv447.
Schoenberg, F. P. (2013), Facilitated estimation of ETAS, *Bull. Seismol. Soc. Am.*, *103*(1), 601–605, doi:10.1785/0120120146.
Schoenberg, F. P., A. Chu, and A. Veen (2010), On the relationship between lower magnitude thresholds and bias in epidemic-type aftershock sequence parameter estimates, *J. Geophys. Res.*, *115*, B04309, doi:10.1029/2009JB006387.
Schurr, B., et al. (2014), Gradual unlocking of plate boundary controlled initiation of the 2014 Iquique earthquake, *Nature*, doi:10.1038/nature13681.
Sornette, D., and A. Helmstetter (2002), Occurrence of finite-time singularities in epidemic models of rupture, earthquakes, and starquakes, *Phys. Rev. Lett.*, *89*(15), 158,501, doi:10.1103/PhysRevLett.89.158501.
Sornette, D., and S. Utkin (2009), Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks, and aftershocks, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, *79*(6 Pt 1), 061110, doi:10.1103/PhysRevE.79.061110.
Sornette, D., and M. J. Werner (2005), Apparent clustering and apparent background earthquakes biased by undetected seismicity, *J. Geophys. Res.*, *110*, B09303, doi:10.1029/2005JB003621.
Turcotte, D. L., J. R. Holliday, and J. B. Rundle (2007), BASS, an alternative to ETAS, *Geophys. Res. Lett.*, *34*, L12303, doi:10.1029/2007GL029696.
Utsu, T. (1961), A statistical study on the occurrence of aftershocks, *Geophys. Mag.*, *30*(4), 521–605.
Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995), The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth*, *43*(1), 1–33, doi:10.4294/jpe1952.43.1.
Van Stiphout, T., J. Zhuang, and D. Marsan (2012), Models and techniques for analyzing seismicity: Seismicity declustering, *Community Online Resour. Stat. Seismol. Anal.*, doi:10.5078/corssa-52382934.
Veen, A., and F. P. Schoenberg (2008), Estimation of space-time branching process models in seismology using an EM-type algorithm, *J. Am. Stat. Assoc.*, *103*(482), 614–624, doi:10.1198/016214508000000148.
Vere-Jones, D. (1970), Stochastic models for earthquake occurrence, *J. R. Stat. Soc.*, *32*(1), 1–62.
Vere-Jones, D. (1978), Earthquake prediction—A statistician's view, *J. Phys. Earth*, *26*(2), 129–146, doi:10.4294/jpe1952.26.129.
Vere-Jones, D. (2005), A class of self-similar random measure, *Adv. Appl. Probab.*, *37*(4), 908–914.
Wang, Q., D. D. Jackson, and J. Zhuang (2010a), Missing links in earthquake clustering models, *Geophys. Res. Lett.*, *37*, L21307, doi:10.1029/2010GL044858.
Wang, Q., F. P. Schoenberg, and D. D. Jackson (2010b), Standard errors of parameter estimates in the ETAS model, *Bull. Seismol. Soc. Am.*, *100*(5A), 1989–2001, doi:10.1785/0120100001.
Werner, M. J. (2008), On the fluctuations of seismicity and uncertainties in earthquake catalogs: Implications and methods for hypothesis testing, PhD thesis, 308 pp., Univ. of Calif.

Werner, M. J., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2011), High-resolution long-term and short-term earthquake forecasts for California, *Bull. Seismol. Soc. Am.*, *101*(4), 1630–1648, doi:10.1785/0120090340.

Zaliapin, I., A. Gabrielov, V. Keilis-Borok, and H. Wong (2008), Clustering analysis of seismicity and aftershock identification, *Phys. Rev. Lett.*, *101*(1), 018501, doi:10.1103/PhysRevLett.101.018501.

Zhuang, J. (2011), Next-day earthquake forecasts for the Japan region generated by the ETAS model, *Earth Planets Space*, *63*(3), 207–216, doi:10.5047/eps.2010.12.010.

Zhuang, J., and S. Touati (2015), Stochastic simulation of earthquake catalogs, *Community Online Resour. Stat. Seismol. Anal.*, doi:10.5078/corssa-43806322.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.*, *97*(458), 369–380, doi:10.1198/016214502760046925.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2004), Analyzing earthquake clustering features by using stochastic reconstruction, *J. Geophys. Res.*, *109*, B05301, doi:10.1029/2003JB002879.

Zhuang, J., D. Harte, M. J. Werner, S. Hainzl, D. Harte, and S. Zhou (2011), Basic models of seismicity: Spatiotemporal models, *Community Online Resour. Stat. Seismol. Anal.*, doi:10.5078/corssa-07487583.